Rapid linkage disequilibrium decay in the Lr10 gene in wild emmer wheat (Triticum dicoccoides) populations

Sela, H ; Loutre, C ; Keller, B ; Nevo, E ; Korol, A ; Fahima, T

Abstract: INTRODUCTION: Recombination is a key evolutionary factor enhancing diversity. However, the effect of recombination on diversity in inbreeding species is expected to be low. To estimate this effect, recombination and diversity patterns of Lr10 gene were studied in natural populations of the inbreeder species, wild emmer wheat (Triticum dicoccoides). Wild emmer wheat is the progenitor of most cultivated wheats and it harbors rich genetic resources for disease resistance. Lr10 is a leaf rust resistance gene encoding three domains: a coiled-coil, nucleotide-binding site, and leucine-rich repeat (CC-NBS-LRR). RESULTS: Lr10 was sequenced from 58 accessions representing 12 diverse habitats in Israel. Diversity analysis revealed a high rate of synonymous and non-synonymous substitutions (d (S) = 0.029, d (N) = 0.018, respectively) in the NBS-LRR domains. Moreover, in contrast to other resistance genes, in Lr10 the CC domain was more diverse than the NBS-LRR domains (d (S) = 0.069 vs. 0.029, d (N) = 0.094 vs. 0.018) and was subjected to positive selection in some of the populations. Seventeen recombination events were detected between haplotypes, especially in the CC domain. Linkage disequilibrium (LD) analysis has shown a rapid decay from r (2) = 0.5 to r (2) = 0.1 within a 2-kb span. CONCLUSION: These results suggest that recombination is a diversifying force for the R-gene, Lr10, in the selfing species T. dicoccoides. This is the first report of a short-range LD decay in wild emmer wheat.

DOI: https://doi.org/10.1007/s00122-010-1434-2

Rapid linkage disequilibrium decay in the *Lr10* gene in wild emmer wheat (*Triticum dicoccoides*) populations

Hanan Sela[1], Caroline Loutre[2], Beat Keller[2], Alan Schulman[3,4], Eviatar Nevo[1], Abraham Korol[1]and Tzion Fahima[1,5]

1.Department of Evolutionary and Environmental Biology, Institute of Evolution, Faculty of Natural Sciences , University of Haifa, Mt. Carmel, Haifa, 31905 Israel

2.Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

3. MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, University of Helsinki, P.O. Box 65, FIN-00014 Helsinki, Finland

4. Biotechnology and Food Research, MTT Agrifood Research Finland, Mylltie 1, FIN-31600 Jokioinen, Finland

5. Corresponding author: Tzion Fahima, Department of Evolutionary and Environmental Biology, Institute of Evolution, Faculty of Natural Sciences , University of Haifa, Mt. Carmel, Haifa, 31905 Israel. Fax: +97248288602, E_mail: fahima@research.haifa.ac.il

Abstract

Recombination is a key evolutionary factor enhancing diversity. However, the effect of recombination on diversity in inbreeding species is expected to be low. To estimate this effect, recombination and diversity patterns of *Lr10* gene were studied in natural populations of the inbreeder species, wild emmer wheat (*Triticum dicoccoides*). Wild emmer wheat is the progenitor of most cultivated wheats and it harbours rich genetic resources for disease resistance. *Lr10* is a leaf rust resistance gene encoding three domains: a coiled coil, nucleotide binding site, and leucine rich repeat (CC-NBS-LRR). *Lr10* was sequenced form 58 accessions from 12 populations representing diverse habitats in Israel. Diversity analysis revealed a high rate of synonymous and non-synonymous substitutions (d_S = 0.029, d_N=0.018, respectively) in the NBS-LRR domains. Moreover, in contrast to other resistance genes, in *Lr10* the CC domain was more diverse than the NBS-LRR domains (d_S = 0.069 vs. 0.029, d_N=0.094 vs. 0.018) and was subjected to positive selection in some of the populations. Seventeen recombination events were detected between haplotypes, especially in the CC domain. Linkage disequilibrium (LD) analysis has shown a rapid decay from r^2=0.5 to r^2=0.1 within a 2 kb span. These results suggest that recombination is a diversifying force of the R-gene, *Lr10*, in the selfing species *T. dicoccoides*. This is the first report of a short range LD decay in wild emmer wheat.

Keywords: gene conversion, *Lr10*, linkage diseqilibrium, Nucleotide diversity, R-genes, recombination, *Triticum dicoccoides*, wild emmer wheat

Introduction

Recombination is a major mechanism underlying genetic diversity (Korol et al. 1994; Gaut et al. 2007; Wicker et al. 2007; Morrell et al. 2006). Tracking recombination events can illuminate the evolution of genes and genomes. Recombination and linkage disequilibrium (LD) are basic tools in genetic mapping and positional cloning of genes in plant genomes (Ersoz et al. 2007; Waugh et al. 2009). Unlike linkage studies of bi-parental mapping populations, in wild populations LD patterns depend on a long history of recombination and therefore enable to obtain improved mapping resolution. On the other hand, LD is affected by population structure and selection, and hence linkage between markers may not necessarily reflect physical proximity (Flint-Garcia et al. 2003; Rostoks et al. 2006). The extent of LD in barley, a selfing plant and a relative of wheat, was found to be much lower in wild populations (*Hordeum spontaneum*) than in a cultivated collection (*H. vulgaris*). LD was rapidly decaying over the first kb in the wild forms while extending over more than 200 kb in elite cultivars (Caldwell et al. 2006). In durum wheat (*Triticum durum*), LD stretches for a long distance (Maccaferri et al. 2005), with high LD values found between loci located 50 cM apart. In bread wheat collections (T. aestivum), LD extent ranged from 1 cM to 5 cM (Chao et al. 2007; Horvath et al. 2009; Somers et al. 2007). In *Aegilops tauschii*, the wild donor of the D genome to *T. aesitvum*, a 1400 bp LD block was observed in the grain softness locus (Massa and Morris 2006). In wild emmer wheat *(T. dicoccoides)* populations from Amiad and Tabihja in northern Israel, high LD values were observed even between loci on different chromosomes. It seems that LD values in *T. dicoccoides* populations are inflated by the population structure (Li et al. 2000a; Li et al. 2000b). To date, no study has targeted the intra-genic LD in *T. dicoccoides* populations and only two studies targeted LD within R-genes in selfers (Mauricio et al. 2003; Kunag et al 2008). Therefore, the current study may add insights into LD patterns in *T. dicoccoides* populations and LD patterns in R-genes.

Leaf rust resistance gene *Lr10* is one of the few cloned disease resistance genes (R-genes) in bread wheat. *Lr10* encodes a three domain protein: coiled- coil, nucleotide binding site and leucine rich

repeats (CC-NBS-LRR) (Feuillet et al. 2003). *Lr10* is located 36 cM from the centromere on the short arm of chromosome 1A. Previous studies have shown that *Lr10* is a very variable (polymorphic) gene while the CC domain is its most diverse part. The high diversity in the CC domain is probably due to positive selection (Loutre et al. 2009). Many plant R-genes exhibit high diversity, positive selection and sequence exchanges since the arms race with pathogens drives them to change recognition specificity frequently (Wicker et al. 2007; Bergelson et al. 2001; Mondragon-Palomino and Gaut 2005). In some R-genes, diversity is maintained by frequency dependent selection where alleles are cycled between high and low frequency in a long co-evolution with pathogens in a form of "trench warfare". The trench warfare model predicts that the polymorphism will be old and under balancing selection (Holub 2001; Stahl et al. 1999). Although many studies of R-gene diversity were previously described (reviewed in Mcdowell and Simon (2006)), most of them were limited to a few alleles or to paralogs and only a few studies focused on population genetics (Mauricio et al. 2003; Meaux and Neema 2003; Butterbach 2007; Yahiaoui et al. 2008; Kuang et al. 2008). These studies mainly targeted the LRR domain, which is considered the most diverse domain due to its interaction with pathogen effectors. However, in *Lr10,* the CC domain was found more diverse than the LRR domain (Loutre et al. 2009). Therefore, in the current study, the CC domain was analyzed separately from the NBS and LRR domains.

T. dicoccoides is the tetraploid (genome BBAA) progenitor of most cultivated wheats. The natural habitats of *T. dicoccoides* are located in the Middle East Fertile Crescent. *T. dicoccoides* populations harbor a valuable pool of resistance genes that can be transferred into cultivated wheat (Rong et al. 2000; Knott et al. 2005; Mergoum et al. 2005; Uauy et al. 2005; Marais et al. 2005; Gustafson et al. 2009). Numerous studies of *T. dicoccoides* genetic diversity have shown that the population in Israel is sub-divided. The selfing nature of emmer wheat and the large difference between *T. dicoccoides* habitats in Israel largely contributes to this pattern (Golenberg and Nevo 1987; Nevo et al. 2002; Peleg et al. 2008b; Sela et al. 2009).

Leaf rust, caused by *Puccinia triticina,* is the most common rust disease of wheat. It is highly

specific and can spread thousands of kilometers by wind (Bolton et al. 2008). In Israel, the pathogen cannot complete its life cycle due to the absence of alternate host and harsh summer conditions. It is assumed that new inocula of *P. triticina* arrive every year from regions adjacent to Israel (Y. Anikster, personal communication). Most of *T. dicoccoides* accessions from Israel are susceptible to leaf rust and only a small fraction of the accessions are resistant (Anikster et al. 2005; Moseman et al. 1985). Nevertheless, leaf rust resistance genes from *T. dicoccoides* were introgressed into cultivated wheat (Marais et al. 2005).

The main objectives of the present study were the following: (i) to estimate *Lr10* nucleotide diversity among emmer wheat populations originating from Israel; (ii) test the differences in *Lr10* nucleotide diversity between and within *T. dicoccoides* populations; (iii) estimate the extent of LD along the gene; and study the effect of recombination on sequence diversity.

Materials and methods

One hundred *T. dicoccoides* accessions from 12 populations in Israel and vicinity were used in the current study (Table 1; Fig 1). A full description of their geographic origin and climatic conditions can be found in Nevo and Beiles (1989). DNA was extracted from 10- day- old seedlings using ArchivePure DNA extraction kit (5 prime). The whole *Lr10* gene (4 kb) was amplified using Pfu Ultra Fusion II HS polymerase (Stratagene) according to the supplier's protocol (annealing temperature was 53ºC and elongation time of 1 min) with the primers ThLR10_V (CGGAACTATGGAGAGTGAAC) and ThLR10_U (GGGAAATGTAGACAGGTACAT) (Feuillet et al. 2003). PCR products were separated on agarose gel and extracted using MinElute gel extraction kit (Qiagene). The extracted DNA was cloned into pSMART vector using GC cloning kit (Lucigen). Samples were sequenced using Big Dye Terminator chemistry on ABI 3700 or ABI 3730 instruments (Applied Biosystems). Alternatively, eight PCR fragments covering the whole gene region were extracted from the gel and sequenced directly (Table 2). Sequence reads were checked and assembled using Phred, Phrap, and Consed software provided by B. Ewing, P. Green & D.

Gordon (available at http://www.phrap.org/). The sequences were aligned using MUSCLE (Edgar 2004) and manually corrected using BioEdit (Hall 2007).

Statistical analysis

DNA sequence alignment was analyzed for nucleotide diversity of synonymous (d_S) and non-synonymous sites (d_N) (Nei 1987) estimated by Kumar's method as implemented in MEGA software (Kumar et al. 2004). Z tests for selection based on the difference between synonymous and non-synonymous substitutions (d_S-d_N) were conducted using MEGA software as well (Nei and Kumar 2000). A consensus maximum parsimony (MP) tree from 500 bootstrapped trees was constructed using MEGA. Pairwise distances between all sequences were calculated using MEGA. Estimation of evolutionary distances were calculated using the maximum composite likelihood (Tamura and Kumar 2002). The rate of variation among sites was modeled with a gamma distribution (shape parameter = 1). The differences in the composition bias among sequences were considered in evolutionary comparisons (Tamura et al. 2004). The evolutionary distances were calculated separately for the CC domain and for the NBS-LRR domains. Based on the distances, a principal coordinates analysis (PCO) was conducted using GENALEX (Peakall and Smouse 2006). Raw precipitation data for each population in the last 20 years was obtained from the Israeli Meteorology Service (Beit-Dagan, Israel) and summarized by SPSS (SPSS Inc.). Monthly and annual means and the variability of rain in the sites of collection of the populations (Table S1) were tested for correlations with diversity indices. Variability of rain was calculated as standard deviation / annual mean.

Linkage disequilibrium (LD) coefficients - r^2 and p-values were calculated in TASSEL (Bradbury et al. 2007). For convenience, p-values were transformed into -log p (-logP) values. Data was filtered for SNPs with more than 20% frequency of the minor allele. In order to minimize the effect of population structure, a subset of 20 sequences that represents major branches on the MP tree was selected (Fig. S2). This subset was analyzed and compared to the whole data set. Regressions and correlations were analyzed in SPSS (SPSS inc.). Recombination events in the sequence alignment were detected with RDP software that combines seven different recombination detection methods (Martin et

al. 2005). Recombination events were confirmed when at least three methods showed $p<0.01$ for the event. The population recombination parameter rho (ρ) (Hudson 2001) and the population mutation parameter theta (θ) (Watterson 1975) were estimated using "rhomap" software in LDhat package (Auton and McVean 2007). All the 33 haplotypes revealed in the alignment were used in the analysis. All polymorphic sites were included for the calculation of theta. Only sites with two alleles were used for the estimation of ρ. The run parameters consisted of 1,100,000 iterations, 100,000 iterations of burn-in period, and sampling interval every 100 iterations. Illegitimate recombination is a process of asymmetric pairing of two dispersed homologous sequences of only a few bp long followed by sequence exchange that can result in either duplications or deletions (Devos et al. 2002). Illegitimate recombination events were detected using Dotter (Sonnhammer and Durbin 1995) and Bioedit (Hall 2007) following the pattern described in Wicker et al. (2007). Two tests were implemented to test for deviation from neutrality of the haplotypes resulted from illegitimate recombination. Slatkin exact test (Slatkin 1994) was conducted using Arelquin software (Excoffier et al. 2005), and Tajima D test (Tajima 1989) was carried out in DNAsp (Rozas et al. 2003)

Resistance tests

Ten day old seedling of all T. dicoccoides accessions were tested for leaf rust resistance as described by Schachermayr et al. (1995) using isolate BRW97512-19, which is avirulent to the *Lr10* resistance gene in bread wheat (*T. aestivum*). Seedlings were sprayed with the rust isolate suspended in Soltrol 170 oil (Chevron Phillips Chemical) and incubated overnight at high humidity (90%) and low temperature (16ºC). Visual scoring of the phenotypes were performed 10 days after infection on a 0-4 scale (0=resistant; 4= susceptible) (McIntosh et al. 1995).

Results

A total of 100 accessions of *T. dicoccoides* from 12 populations were tested for the presence of *Lr10* (Fig 1). In 95 accessions *Lr10* was present as determined by PCR amplification of at least three primer pairs listed in Table 2. The full length 4 kb *Lr10* from 58 accessions was cloned and sequenced. First, reads were assembled into contigs for each accession and then DNA and protein sequence alignments were generated. The sequences were deposited to GenBank (accsesions GU393247-GU393304). Four accessions, originated from two populations (Jaba and Givat- Koach), had a 1.2 kb deletion of the NBS domain and therefore were excluded from the diversity analysis. Three sequences obtained from Mt. Hermon accessions had a premature stop codon at position 130.

Sequence diversity

Out of 54 *Lr10* sequences, 33 different haplotypes were detected in the alignment, 332 sites out of 4114 sites were polymorphic, and the nucleotide diversity (π) was 0.029. Within-population nucleotide diversity for coding regions was calculated for nine populations that were represented by at least three accessions in the alignment. Previous studies have shown that the CC domain is the most diverse region of the *Lr10* (Feuillet et al. 2003; Loutre et al. 2009). Therefore, the analysis was conducted on the CC domain and on the NBS-LRR domains separately. Diversity values for synonymous and non- synonymous substitution rate (d_S and d_N, respectively) of the NBS-LRR domains differed significantly between populations (p= 0.04) (Table 3). The highest diversity was observed in the population from Kokhav-Hashahar (d_S=0.036, d_N=0.021) while the lowest diversity was observed in the population from Beit-Oren (d_S=0, d_N=0; Table 3). Overall d_S value was 0.0295 while d_N value was only 0.0182. The difference d_S-d_N was mainly positive and varied between populations, ranging from -0.0003 in Gamla population to 0.0154 in Kokhav-Hashahar population. Z tests for purifying selection based on d_S-d_N values revealed significant purifying selection in the populations of Tabigha, Mt. Gilboa, Gitit, and Kokhav-Hashahar ($p<0.05$) (Table 3). d_S values of these populations were higher than the values of the other populations (p=0.016). The mean d_S value of these populations was significantly higher than mean dS value of the rest of the populations (d_S=0.030 and 0.008 respectively, p=0.016), while no significant difference was

observed for d_N values (d_N=0.017 and 0.006 respectively; p=0.11). The population diversity values, d_S and d_N, were tested for correlation with previously obtained Simple Sequence Repeat markers (SSR) gene diversity (*He*) values of the same populations (Peleg et al. 2008b) (Table S1). d_S and d_N were positively correlated with *He* values (r=0.73, p= 0.027 and r=0.76, p=0.021, respectively) (Fig 2 A,B). In the CC domain, mean population d_S and d_N values were positively correlated with the NBS-LRR d_S and d_N values (r=0.81 p=0.008 and r=0.80, p=0.048, respectively). The overall d_S value of the CC domain was 0.069, which is more than two fold higher than the NBS-LRR d_S value (0.0295), and the overall d_N value for the CC domain was 0.094, which is five fold higher than NBS-LRR d_N value (0.0182). In contrast to NBS-LRR values, d_S-d_N values of the CC domain were negative. The overall d_S-d_N value was -0.025 and ranged among the populations from -0.05 in Kokhav-Hashahar population to 0 in Beit-Oren population. In four populations (Mt. Hermon, Mt. Gilboa, Kokhav-Hashahar and Amirim) a significant positive selection was detected.

Climatic factors associated with diversity

Searching for climatic factors that are associated with diversity and selection patterns have revealed that populations with significant d_S-d_N>0, suggesting purifying selection in the NBS-LRR domains, were collected from sites with low annual precipitation and high variation of annual rain. In the sites where those populations were collected, the annual precipitation ranged from 375 mm to 432 mm and rain variation ranged from 0.34 to 0.46, while in the other sites the ranges were 535-1300 mm and 0.29-0.32, respectively (Table S1). NBS-LRR d_S and d_S-d_N values of the populations were positively correlated with rain variation in the collection sites (Spearman rank correlation r_s=0.79, p=0.011 and r_s=0.84, p=0.005, respectively) (Fig 2.C,D).

Distributions of haplotypes:

The 33 haplotypes were not evenly distributed. Two haplotypes with 99.8% identity were present in 14 out of the 58 accessions (24%). The two haplotypes together were fixed in the populations of Gamla, Rosh-Pinna and Beit-Oren (30% of the populations). Another haplotype was present in 6

accessions, 4 of them from Amirim population and another haplotype was present in 5 accessions. All the rest of the haplotypes were present in 3 or less accessions. In 3 populations, Mt. Hermon, Amirim and Gilboa, one haplotype, different in each population, was present in 50% or more of the accessions.

Principal coordinates analysis PCO

Based on the sequence distances of the NBS-LRR domains and the CC domain, two PCO graphs were plotted (Fig 3A&B respectively). Sequences that had the same coordinates were considered as one haplotype. In the plots, many haplotypes were represented by two or more genotypes of the same population and/or genotypes from more than one population. Nevertheless, in most of the cases, populations represented by the same haplotype did not share similar geographic or climatic characteristics. Moreover, different haplotypes of the same population were scattered on the plot and not clustered together. Four *T. urartu* (Loutre et al. 2009) sequences that were included in the plot did not cluster together. The PCO plots of the CC and the NBS-LRR domains showed different patterns. In the CC plot, haplotypes were scattered all over the plot with some clustering but with no obvious eco-geographic relations between the clustered populations. In the NBS-LRR plot, a clear division to two groups was observed. Seven populations and the *T. urartu* accessions were represented in both groups, while four populations were represented by only two haplotypes that were almost identical. The main difference in the two groups of sequences lies in the LRR domain where two well conserved, ~400bp long, haplotypes exist. These haplotypes are highly diverged one from the other. Tajima's test (Tajima 1989) has shown that in a stretch of 114bp sequence within these haplotypes there is a deviation from neutrality suggesting that balancing selection was involved in shaping the diversity ($D= 2.46$, $p<0.05$).

Linkage disequilibrium analysis

Previous analysis of population structure showed that *T. dicoccoides* populations are well structured (Peleg et al. 2008a). Therefore, in order to avoid the bias of the population structure on LD, a subset of 20 sequences was selected from the MP phylogenetic tree (Fig S2). LD values (-logP and r^2)

between all pairwise comparisons of polymorphic sites were plotted against the physical distance (bp) between sites in the whole set and in the selected subset (Fig 4). For the whole set, locally weighted scatterplot smoothing (LOESS) curves of the association values, -logP and r^2, were declining within 3 kb and 2 kb, respectively (Fig 4. A,B). For the subset, the decline was much steeper with sharp LD decay found within only 1kb. The differences in LD decay patterns between the whole set and the subset were more profound for -logP values than for r^2 values. Curve estimation of regression models showed that the logarithmic model is the best one to describe the relationships between LD values and physical distance. LD-r^2 values of the whole set and the subset were correlated with distance (R^2 = 0.40 and R^2= 0.41, respectively, $p<0.001$), while -logP values were correlated for the subset (R^2 = 0.40, $p<0.001$) but much less for the whole set (R^2 = 0.16, $p<0.001$). LD plots along the gene showed the same pattern (Fig. 5). In the plot of the whole set, low p-values were found away from the diagonal line indicating associations not arising from physical proximity. This pattern was not noted neither for p-values of the subset nor for r^2 values of the whole set and the subset. In these plots, small LD blocks (< 250 bp) were observed along the diagonal with the exception of a larger block of ~500 bp at the 3' end of the gene. This block lies in the LRR domain in the region were the two well conserved haplotypes mentioned earlier were observed.

Recombination detection analysis implemented in RDP revealed 17 highly significant recombination events (confirmed by at least three detection methods at the level $p< 0.01$). These events resulted in 22 different combinations of recombinant fragments (Fig 6). The distribution pattern of breakpoints was different between the CC domain and other domains. While in the CC domain most of the break points were found within the domain, in the NBS and LRR domains most of the breakpoints were located near the borders of the domain. To test the significance of this observation, the distance of the breakpoints to the nearest border was calculated and transformed to a relative distance as a fraction of the domain size. The distances in the CC domain were significantly higher than in the NBS-LRR domains (mean 0.6 and 0.25, respectively; p=0.004). A

significant difference (p=0.001) was found in the length of the recombinant fragments between the domains. In the CC domain fragments were short, ranging from 23 to 361 bp (mean 177 bp) while in the NBS and LRR domains fragments were longer, ranging from 349 to 3020 bp (mean 1246 bp). Only one breakpoint was found in the intron region. All events occurred between sequences obtained from different populations and were detected in 2 to 30 pairs of sequences (mean 11). The estimation of population recombination parameter rho (ρ) was 0.012 per base pair and the estimation of population mutation parameter, theta (θ), was 0.021 per base pair. The ratio between recombination and mutation rates (ρ /θ) was 0.59.

Ten illegitimate recombination sites were detected in the alignment. Three events have resulted in deletions of 20-26 bp in the intron region and four more resulted in tri-nucleotide deletions that were scattered along the gene. These deletions were present in 2-32 of the sequences. One repeat of 81 bp in the CC domain was found only in one sequence. Another repeat of 32 bp was found in the intron region of all sequences. This repeat was about 40% degenerated and a subsequent deletion was nested within it indicating that this region is prone to illegitimate recombination (Fig 7A,B). None of the illegitimate recombination events observed in the alignment had caused frame shift. A region with high indel variation was revealed between the CC and the NBS domains (positions 380 to 460) where six indel haplotyps were observed (Fig.7C). A repeat of 24 bp in the region, caused by illegitimate recombination, and subsequent recombination errors formed a 54 bp difference between the shortest and the longest sequences. However, none of these indels have caused any frame shift. This region is a linker between the CC and the NBS domains. Slatkin's exact test and Tajima's D test did not reveal a significant deviation from neutrality in this region.

Resistance tests

Resistance tests were conducted with the leaf rust isolate BRW97512-19, which is avirulent to some *Lr10* allels from *T. durum* and *T. aestivum*. All 100 wild emmer wheat accessions tested in the current study showed high infection types (3 to 4) with this isolate.

Discussion

In the current study, the sequence diversity of *Lr10* within and between *T. dicoccoides* populations from Israel was estimated, the LD and recombination patterns along the *Lr10* gene were analysed, and the processes shaping the diversity and the LD were revealed.

Lr10 diversity within and between populations

The diversity revealed in the current paper is of the same magnitude and distribution of synonymous and nonsynonymous substitutions along the gene those observed by Loutre et al. (2009). *T. dicoccoides* populations showed large differences in diversity. These differences could be caused by the population demographic history, e.g. bottlenecks, or by selection pressure of the pathogen. The d_S and d_N values of the NBS-LRR domains obtained in this study were strongly correlated with the diversity index *He* obtained for SSR markers in the same populations studied earlier (Peleg et al. 2008b). This correlation suggests that demographic processes have a large effect on Lr10 nucleotide variation, since they act similarly on all loci. However, the negative d_S-d_N values in the CC domain obtained in all populations indicate that this domain is under positive selection. Such positive selection in the CC domain might indicate this domain interact with pathogen effectors as suggested by Loutre et al. (2009) for *Lr10* and by Bergelson et al. (2001) in general. d_S-d_N values differed several fold between populations both in the CC domain and in the NBS-LRR domains. In the NBS-LRR domains, a significant d_S-d_N>0 difference in four populations was observed suggesting the involvement of purifying selection. However, the significant d_S-d_N>0 values in the populations resulted from high d_S values and not from low d_N values. Therefore, the results may not reveal higher purifying selective pressure in the populations but rather higher diversity in the populations. Nevertheless, purifying selection may have acted on the NBS-LRR domains and it's signature was maintained in these populations while decaying in the other populations. The populations with significant d_S-d_N>0 values originated in the transition zone between Mediterranean and steppe climate with annual precipitation of less than 450 mm and rain variability higher than 34%. The high d_S values may reflect the overall high genomic variability in

this climatic transition zone as suggested by Safriel et al. (1994), Volis et al. (2001), Kark et al. (2005), Kark et al. (2008) and Peleg et al. (2008b). In the CC domain, no clear pattern was observed for the distribution of populations with significant d_S-d_N<0 values.

The frequency of the haplotypes and their distribution was not even. Two very similar haplotypes accounted for 24% of the accessions and were fixed in 3 populations. The Beit Oren population is distant from Gamla and Rosh-Pinna populations and does not share with them climatic characteristics. Therefore, the reason for the fixed haplotypes in these populations is not clear. It is possible that these haplotypes have some selective advantage raising its frequency in the populations but why in these populations and not in others?. Alternatively, genetic drift can explain fixation but then why the same haplotypes are fixed in three different populations? These haplotypes should be prioritized for resistance tests with different races of leaf rust as they may be more resistant than other haplotypes.

The PCO analysis used in the current study showed that the revealed diversity is old and probably existed before populations were separated to the current status, since haplotypes from the same populations did not form clusters. Furthermore, the haplotypes of *T. urartu,* the donor of the A genome to *T. dicoccoides,* are scattered and are not clustered. This pattern was also observed in Loutre et al. (2009) where different *Triticum* species shared the same *Lr10* haplotypes. The different pattern observed between the CC PCO and the NBS-LRR PCO could point to different selection pressures and/or high recombination rate between the domains. The old origin of diversity is supported by the observation that most recombination events were detected between sequences obtained from different populations, implying that the resulting haplotypes were formed before populations separated. Significant gene flow could also contribute to low population differentiation but then one would expect geographical relationships between clustered accessions and this is not the case. Moreover, *T. dicoccoides* is a selfer with low gene flow rate (Golenberg and Nevo 1987). The genetic distance between *T. dicoccoides* populations is not correlated with geographic distance

in contrary to the expectations for situations where gene flow is substantial (Peleg et al. 2008b; Sela et al. 2009; Haudry et al. 2007). The ancient diversity observed in the current study supports a long evolutionary history of *Lr10* - pathogen effector interaction in the "trench warfare" model where polymorphism is maintained by alleles that are recycling between high and low frequencies shifted by frequency dependent selection (Holub 2001; Stahl et al. 1999). The long coalescent time is also supported by the significant Tajima test of the two haplotyps in the LRR domain suggesting balancing selection. These findings of old polymorphism are in agreement with the observations of Loutre et al. (2009).

Resistance tests with a leaf rust isolate which is avirulent to *Lr10* of bread wheat did not detect any *Lr10*-dependent resistance in the collection of *T. dicoccoides* tested in the current study. Therefore, it was impossible to assess the direct impact of mutations of the different haplotypes on gene function and specificity. However, in most accessions, an intact open-reading frame was observed and the gene was expressed in all eight accessions that were tested for *Lr10* mRNA (data not shown). These findings suggest that *T. dicoccoides Lr10* haplotypes are functional against unknown leaf rust isolates present now, or that were present in the recent past.

Recombination and linkage disequilibrium

The alignment of *Lr10* sequences obtained from 54 accessions of *T. dicoccoides* revealed many recombination events. In the CC domain there were mostly very short recombinant sequences that are typical of gene conversion (Jeffrys and May 2004), while the NBS and LRR domains exhibited longer sequence exchanges that could result from crossing over events. These results are in agreement with LD analysis, which showed an LD block at the 3' end (Fig. 5). Recombination events were detected between *Lr10* sequences that were derived from different *T. dicoccoides* populations and not within populations. This may suggest that recombination events took place prior to the current separation between *T. dicoccoides* populations. Old recombination events imply that selective pressures may have an effect on the recombination outcome. Shuffling haplotypes at

the CC domain enhances diversity and may have a selective advantage if this domain is involved in pathogen recognition as predicted. In the LRR domain, two very conserved haplotypes were observed. These two haplotypes have a long evolutionary history and can be tracked back to T. *urartu,* the A genome donor of *T. dicoccoides* (Loutre et al. 2009). It seems that recombination is suppressed in this region. Apparently, there is a tendency not to shuffle haplotypes within the NBS and LRR domains, since most of the breakpoints in the region are restricted to the domain borders. (Michelmore and Meyers 1998) expected that inter-allelic recombination is enhancing diversity in the LRR domain of R-genes. This expectation was validated by Kuang et al. (2004) and Yahiaoui et al. (2006). The data presented here about the LD block in the *Lr10* LRR domain does not match this expectation but in the CC domain this expectation can be met. The LD block in the *Lr10* LRR domain is similar to the LD block that was observed in the *Rps2* LRR domain from *Arabidopsis thaliana,* both blocks are located around a hot spot of diversity. The *Rps2* hot spot of divergence differentiate between resistant and susceptible clades while the *Lr10* hot spot does not distinguish between resistant and susceptible haplotypes (Mauricio et al. 2003; Loutre et al. 2009). It could be because the *A. thaliana Rps2* block is located in the 5' region of the LRR domain while in the current *Lr10* study the LD block is located in the 3' region. The two parts of the LRR play a different role in the protein function (Lukasik and Takken 2009).

Understanding LD decay patterns is essential when conducting association studies or for estimating rates of recombination. Population structure can lead to spurious LD associations especially in selfing species (Ewens and Spielman 1995). In the current study r^2 values were better predictors of physical distance than LD p-values since they were less affected by population structure. Choosing a subset of sequences from a tree proved to be a good method to avoid population structure bias. This approach is somewhat similar to the method implemented by Breseghello and Sorrells (2006), even though they have removed only very similar haplotypes and in the current study only distant haplotypes were selected. Taking into account that *T. dicoccoides* is a selfer, LD decay in *Lr10* was very rapid, within 1-2 kb. This LD extent was also found in wild barley, *Hordeum spontaneum,* a

relative of wheat and a selfer as well (Morrell et al. 2006; Caldwell et al. 2006). Whole genome LD studies in the selfer *A. thaliana* have shown a longer extent of LD, up to 250 kb (Nordborg et al. 2002). LD in bread wheat collections extended 1-5 cM (Chao et al. 2007; Horvath et al. 2009; Somers et al. 2007). Longer LD extent may be due to longer stretches between the sampled loci that overlooks local gene conversion events and only displays crossing over events (Plagnol et al. 2006). This situation creates a paradox, where high resolution analysis, in a base pair scale, shows steep decay in LD, while in low resolution analysis, in centiMorgan scale, the decay is by far more moderate (Andolfatto and Nordborg 1998). From a practical point of view, attempts to associate very closely linked SNPs with a trait, especially in the highly diverse R genes, may be hampered by the short LD extent. The rapid decay may be a result of a high number of point mutations or recombination events. In the current study the ratio ρ/θ was lower than 1 (0.59), which may imply that recombination events had a smaller effect than mutations on sequence diversity. Nevertheless, the ρ/θ ratio is fairly high for inbreeders. For reference, the average ρ/θ ratio in inbreeders ranges from 0.05 in *A. thaliana* to 1.5 in wild barley (Morrell et al. 2006; Nordborg et al. 2002). In the outbreeder wild maize, teosinte (*Zea mays ssp. parviglumis*), ρ/θ ratio is 4.5 (Wright et al. 2005).

Recombination is a major force enhancing indel diversity as revealed by the numerous events of illegitimate recombination, especially in the intron region of *Lr10*. One interesting example is the linker region between the NBS and the CC domains of *Lr10* where high length variability was observed between six haplotypes. It is not clear if the variability in this region affects the function of the protein. Several studies have shown that the CC domain in R genes can function even when expressed separately from the rest of the protein (e.g. Rairdan et al. (2008) and Moffett et al. (2002)). Neutrality tests looking for old polymorphism (Slatkin 1994; Tajima 1989) did not reject the neutrality of indel variation of this region but the long and short haplotypes found in all ploidy levels (*T. urartu* AA, *T. dicoccodes* BBAA, *T. aestivum* BBAADD) may indicate that these haplotypes are old and stable (Loutre et al 2009) .

The results presented here highlight the role of recombination in enhancing diversity in wild populations of the selfing plant species *T. dicoccoides.* The results support the trench war-fare model of old polymorphism in R-genes. The high diversity observed in *Lr10* in *T. dicoccoides* populations emphasizes the need to further explore and conserve this diverse gene pool.

Acknowledgments

This work was supported by grants from the European Union's Sixth Framework Programme (FP6) in the BioExploit project (No. CT-2005-513949). we also acknowledge The Israel Science Foundation equipment grants #048/99, #1478/04, #1719/08. The authors wish to thank Ortal Mergi for technical assistance.

References

Andolfatto P, Nordborg M (1998) The Effect of Gene Conversion on Intralocus Associations. Genetics 148:1397-1399

Anikster Y, Manisterski J, Long DL, Leonard KJ (2005) Leaf rust and stem rust resistance in triticum dicoccoides populations in israel. Plant Dis 89:55-62

Auton A, McVean G (2007) Recombination rate estimation in the presence of hotspots. Genome Res 17:1219

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-Genes. Science 292:2281-2285

Bolton MD, Kolmer JA, Garvin DF (2008) Wheat leaf rust caused by *Puccinia triticina*. Mol Plant Pathol 9:563-575

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL:

software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165-1177

Butterbach P (2007) Molecular evolution of the disease resistance gene Rx in Solanum. PhD dissertation. Wageningen University, The Netherlands.

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557-567

Chao S, Zhang W, Dubcovsky J, Sorrells M (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. Wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci 47:1018-1030

Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075-1079

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792

Ersoz ES, Yu J, Buckler ES (2008) Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement. Springer Netherlands, pp 97-119

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57:455

Excoffier L, Laval G, Schneider S (2005) arlequin version 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci U S A 100:15253-15258

Flint-Garcia SA, Thornsberry JM, S E, IV B (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007) Opinion: Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet 8:77-84

Golenberg EM, Nevo E (1987) Multilocus differentiation and population structure in a selfer, wild emmer wheat. *Triticum dicoccoides* Heredity 451–456

Gustafson JP, Raskina O, Ma X, Nevo E (2009). Evolutionary genomics of wheat. In: Carver BF (ed) *Wheat: Science and Trade*. Hoboken, NJ. Wiley-Blackwell. pp. 5-30.

Hall T (2007) BioEdit, version 7.0. 9. Carlsbad, CA: Computer Program and documentation. Ibis Biosciences

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S (2007) Grinding up wheat: A massive loss of nucleotide diversity since domestication. Mol Biol Evol 24:1506-1517

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2:516-527

Horvath A, Didier A, Koenig J, Exbrayat F, Charmet G, Balfourier F (2009) Analysis of diversity

and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 119:1523-1537

Hudson RR (2001) Two locus sampling distributions and their application. Genetics 159:1805-1817

Jeffrys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat Genet 36:151-156

Kark S, Hadany L, Safriel UN, Noy-Meir I, Eldredge N, Tabarroni C, Randi E (2008) How does genetic diversity change towards the range periphery? An empirical and theoretical test. Evol Ecol Res 10:391-414

Kark S, Volis S, Novoplansky A (2004) Biodiversity along core-periphery distribution clines. In: Shachak M, Gosz JR, Pickett STA, Perevolotsky A (eds) Biodiversity in drylands. Oxford, Oxford University Press, pp 30–56

Knott DR, DaPeng B, Zale J (2005) The transfer of leaf and stem rust resistance from wild emmer wheats to durum and common wheat. Can J Plant Sci 85:49-57

Korol AB, Preygel IA, Preygel SI (1994) Recombination variability and evolution. Chapman & Hall, London p 361

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870-2894

Kuang H, van Eck HJ, Sicard D, Michelmore R, Nevo E (2008) Evolution and genetic population structure of prickly lettuce (*Lactuca serriola*) and Its RGC2 resistance gene cluster. Genetics 178:1547-1558

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary

Genetics Analysis and sequence alignment. Brief Bioinform 5:150-163

Li YC, Fahima T, Peng JH, Röder MS, Kirzhner VM, Beiles A, Korol AB, Nevo E (2000a) Edaphic microsatellite DNA divergence in wild emmer wheat, *Triticum dicoccoides*, at a microsite: Tabigha, Israel. TAG Theoretical and Applied Genetics 101:1029-1038

Li Y-, Röder MS, Fahima T, Kirzhner VM, Beiles A, Korol AB, Nevo E (2000b) Natural selection causing microsatellite divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. Theor Appl Genet 100:985-999

Loutre C, Wicker T, Travella S, Galli P, Scofield S, Fahima T, Feuillet C, Keller B (2009) Two different CC-NBS-LRR genes are required for Lr10-mediated leaf rust resistance in tetraploid and hexaploid wheat. Plant J 60:1043-1054

Lukasik E, Takken FLW (2009) STANDing strong, resistance proteins instigators of plant defence. Curr Opin Plant Biol 12:427–436

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271-290

Marais GF, Pretorius ZA, Wellings CR, McCallum B, Marais AS (2005) Leaf rust and stripe rust resistance genes transferred to common wheat from *Triticum dicoccoides*. Euphytica 143:115-123

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21:260-262

Massa A, Morris C (2006) Molecular Evolution of the Puroindoline-a, Puroindoline-b, and Grain Softness Protein-1 Genes in the Tribe *Triticeae*. J Mol Evol 63:526-536

Mauricio R, Stahl EA, Korves T, Tian D, Kreitman M, Bergelson J (2003) Natural selection for

polymorphism in the disease resistance gene Rps2 of *Arabidopsis thaliana*. Genetics 163:735-746

Mcdowell JM, Simon SA (2006) Recent insights into *R* gene evolution. Mol Plant Pathol 7:437-448

McIntosh RA, Wellings CR, Park RF (1995) Wheat rusts: an atlas of resistance genes. CSIRO, East Melbourne, Australia

Meaux J, Neema C (2003) Spatial patterns of diversity at the putative recognition domain of resistance gene candidates in wild bean populations. J Mol Evol 57:S90-S102

Mergoum M, Frohberg RC, Miller JD, Stack RW (2005) Registration of 'Steele-ND'Wheat Registration by CSSA. Crop Sci 45:1163-1164

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113-1130

Moffett P, Farnham G, Peart J, Baulcombe DC (2002) Interaction between domains of a plant NBS−LRR protein in disease resistance-related cell death. EMBO J 21:4511-4519

Mondragon-Palomino M, Gaut BS (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. Mol Biol Evol 22:2444-2456

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2006) Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics 173:1705

Moseman JG, Nevo E, Gerechter-Amitai ZK, El-Morshidy MA, Zohary D (1985) Resistance of *Triticum dicoccoides* collected in Israel to infection with *Puccinia recondita tritici*. Crop Sci 25:262-265

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, USA

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, USA

Nevo E, Korol AB, Beiles A, Fahima T (2002) Evolution of wild emmer and wheat improvement: Population genetics, genetic resources, and genome organization of wheat's projenitor, *Triticum dicoccoides*. Springer

Nevo E, Beiles A (1989) Genetic diversity of wild emmer wheat in Israel and Turkey - structure, evolution, and application in breeding. Theor Appl Genet 77:421-455

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190-193

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288-295

Peleg Z, Fahima T, Abbo S, Krugman T, Saranga Y (2008a) Genetic structure of wild emmer wheat populations as reflected by transcribed versus anonymous SSR markers. Genome 51:187-195

Peleg Z, Saranga Y, Krugman T, Abbo S, Nevo E, Fahima T (2008b) Allelic diversity associated with aridity gradient in wild emmer wheat populations. Plant, Cell Environ 31:39-49

Plagnol V, Padhukasahasram B, Wall JD, Marjoram P, Nordborg M (2006) Relative influences of crossing over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. Genetics 172:2441-2448

Rairdan GJ, Collier SM, Sacco MA, Baldwin TT, Boettrich T, Moffett P (2008) The coiled-coil and nucleotide binding domains of the potato rx disease resistance protein function in pathogen recognition and signaling. Plant Cell 20:739-751

Rong J, Millet E, Manisterski J, Feldman M (2000) A new powdery mildew resistance gene:

Introgression from wild emmer into common wheat and RFLP-based mapping. Euphytica 115:121-126

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci U S A 103:18656-18661

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497

Safriel UN, Volis S, Karr S (1994) Core and peripheral populations and global climate. Isr J Plant Sci 42:331-345

Schachermayr GM, Messmer MM, Feuillet C, Winzeler H, Winzeler M, Keller B (1995) Identification of molecular markers linked to the *Agropyron elongatum*-derived leaf rust resistance gene *Lr24* in wheat. TAG Theor Appl Genet 90:982-990

Sela H, Cheng J, Jun Y, Nevo E, Fahima T (2009) Divergent diversity patterns of NBS and LRR domains of resistance gene analogs in wild emmer wheat populations. Genome 52:557-565

Slatkin M (1994) An exact test for neutrality based on the Ewens sampling distribution. Genet Res 64:71-74

Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557-567

Sonnhammer ELL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited

for genomic DNA and protein sequence analysis. Gene 167:

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400:667-671

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101:11030

Tamura K, Kumar S (2002) Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol 19:1727

Uauy C, Brevis JC, Chen X, Khan I, Jackson L, Chicaiza O, Distelfeld A, Fahima T, Dubcovsky J (2005) High-temperature adult-plant (HTAP) stripe rust resistance gene *Yr36* from *Triticum turgidum ssp. dicoccoides* is closely linked to the grain protein content locus Gpc-B1. Theor Appl Genet 112:97-105

Volis S, Yakubov B, Shulgina I, Ward D, Zur V, Mendlinger S (2001) Tests for adaptive RAPD variation in population genetic structure of wild barley, *Hordeum spontaneum* Koch. Biol J Linn Soc 74:289-303

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256-276

Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. Curr Opin Plant Biol 12:218-222

Wicker T, Yahiaoui N, Keller B (2007) Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. Plant J 51:631-641

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310

Yahiaoui N, Brunner S, Keller B (2006) Rapid generation of new powdery mildew resistance genes after wheat domestication. Plant J 47:85-98

Yahiaoui N, Kaur N, Keller B (2008) Independent evolution of functional *Pm3* resistance genes in wild tetraploid wheat and domesticated bread wheat. Plant J

TABELS

Table 1: List of wild emmer wheat accessions used in the study

Population	Population No[1]	No of Accessions	Accessions
Mt. Hermon	1	6	1-22, 1-24, 1-42, 1-51, 1-59, 1-64
Gamla	8	4	8-22, 8-28, 8-35, 8-40
Rosh Pinna	9	5	9-29, 9-3, 9-33, 9-45, 9-55
Amiad	10	1	10-85
Tabigha	15	3	13-53, 15-25, 15-6
Mt. Giloboa	16	4	16-28, 16-33, 16-34, 16-51
Gitit	18	8	18-1, 18-16, 18-20, 18-25, 18-27, 18-39, 18-4, 18-44
Kokhav-Hashahar	19	4	19-10, 19-13, 19-16, 19-5
Jaba	23	4	23-24, 23-3, 23-45, 23-6,
Amirim	24	5	24-12, 24-13, 24-36, 24-46, 24-47
Achihud	26	1	26-22
Bat-Shelomo	30	2	30-18, 30-4
Beit-Oren	28	5	28-14, 28-2, 28-30, 28-33, 28-53
Givat-Koach	33	4	33-58, 33-13, 33-19, 33-48
Burj el Maliech[2]		1	Be-1
Ma'ale Merav[2]		1	Me-2
T. urartu[3]		4	TU17119, TU17122, TU17159, TU6735
T. aestivum[4]		1	AY2701574

1. Numbers are acording to Nevo and Beiles (1989)
2. *T. dicoccoides* accessions from Peleg *et al*. (2005)
3. GenBank accessions from Loutre *et al*. (2002)
4. GenBank accession from Feuillet *et al*. (2003)

Table.2. Table2. Primers used for amplification and sequncing of overlapping fragments of *Lr10* sequence. Start position is relative to start codon.

Primer name		Size (bp)	Start position
LR10A2F LR10A4R	ATGCATCTTTACCCGTCCAC TGAGCGTGGAATTGAGACAG	840	-100
LR10S1F LR10S1R	GACGACTCTTGCCAAACAGG TGTCAAATTGCCTTTCATCG	831	610
LR10b2F LR10b2R	CTACGTCCGTTATTCAAAAGAT ACCAAATACTCGGTTCAAAA	1500	844
LR10T1F LR10T2R	TGGGATGTTTTGAAATGTGC AGAAGCACCGAGAGATCGAG	800	2024
LR10U1F LR10U1R	GGCACGGTTCTACAGAGTCC AAGCAGCCATTTGTTCATCC	836	2629
LR10V2F LR10V2R	ATGAGCTGGGCAAATTAACG GCCATTTTGTCACTCATCCTG	750	3195

Table 3. Nucleotide diversity values d_N and d_S of *T. dicoccoides* populations

Pop No[1]	Population	N	No of haplo-types	NBS-LRR				CC			
				d_S	d_N	d_S-d_N	Sig [2]	d_S	d_N	d_S-d_N	Sig [3]
1	Mt. Hermon	6	4	0.0262	0.0202	0.003		0.0400	0.0650	-0.030	*
8	Gamla	4	2	0.0000	0.0003	-0.0003		0.0000	0.0000	0.000	
9	Rosh-Pinna	5	2	0.0124	0.0111	0.001		0.0390	0.0450	-0.006	
15	Tabigha	3	3	0.0342	0.0189	0.015	*	0.0540	0.0630	-0.008	
16	Mt. Gilboa	4	2	0.0240	0.0137	0.010	*	0.0310	0.0650	-0.030	*
18	Gitit	8	4	0.0291	0.0173	0.011	*	0.0670	0.0820	-0.018	
19	Kokhav-Hashahar	4	3	0.0363	0.0210	0.015	*	0.0560	0.1040	-0.050	*
24	Amirim	5	2	0.0029	0.0015	0.001		0.0400	0.0740	-0.030	*
28	Beit-Oren	5	1	0.0000	0.0000	0.000		0.0000	0.0000	0.000	
	Mean			0.0230	0.0200	0.010		0.0400	0.0570	-0.017	
	Overall	44	15	0.0295	0.0182	0.010	*	0.0690	0.0940	-0.025	

1. Numbers of populations are according to Nevo and Belis (1989).

2.NBS-LRR Sig: populations with significant purifying selection in the NBS-LRR domains ($p<0.05$).

3. CC Sig: populations with significant positive selection in the CC domain ($p<0.05$).

FIGURES

Fig 1: Geographic distribution of the 12 populations of wild emmer wheat, *T. dicoccoides*, tested in this study. For names of the numbered populations see list in Table 1. The numbers of populations are according to Nevo and Beiles (1989).

Fig 2: Correlations of diversity indices and rain variability. A. Correlation of d_N values of the populations with SSR *He* (gene diversity) of the same populations (PELEG *et al.* 2008b). B. Correlation of d_S values of the populations with SSR *He*. C. Correlation of d_S-d_N values of the populations with rain variability (SD of annual precipitation/ mean annual precipitation). D. Correlation of d_S values of the populations with rain variability.

Fig 3: Principal coordinates analysis (PCO) based on the genetic distances between *Lr10* sequences of *T. dicoccoides* accessions from Israel. **A.** PCO based on NBS-LRR domains. **B.** PCO based on CC domain. Diamonds represent position of haplotypes. Numbers near the diamonds are the populations represented in each haplotype x number of accessions represented. Population numbers are according to Table 1. TA is ThatcherLr10, the first cloned *Lr10* from *T. aestivum* cv. *ThatcherLr10* (Feuillet *et al*. 2003). TU are *T. urartu* accessions (Table 1). Arrows connect the most frequent haplotypes to their position.

Fig 4: LD values of pairwise comparisons between polymorphic sites plotted against their physical distances (bp). LOESS curves are in blue**,** logarithmic regression curves are in red. **A, B:** -log P and r^2 values calculated for the whole dataset (60 accessions), respectively. **C, D:** -log P and r^2 values of the subset (20 selected accessions), respectively.

Fig 5: LD plot of *p* and r^2 values of pairwise comparisons between polymorphic sites. **A.** Values for the whole set. **B.** Values for the subset. Upper triangles are r^2 values; lower triangles are p-values. Only polymorphic sites (f>0.2) are represented.

Fig 6: Schematic sequence display of recombination events detected by RDP (Martin *et al*. 2005).

Full length bars are the acceptor (“daughter” in RDP terms) sequence. Short bars below each full length bar are recombinant fragments (“parent” in RDP terms). Different colors represent different representative accessions. Vertical lines are the domains borders.

Fig 7: Illegitimate recombination and indel diversity. A. A typical deletion caused by illegitimate recombination. The upper sequence consists of a tri-nucleotide repeat (shaded). The sequence between the tri-nucleotide repeats was deleted in the lower sequence. Sequence matches are shaded.

B. An example of a degenerated repeat present in all sequences. A sequence aligned against itself. Bold letters are the short hepta-nucleotide repeat that is the cause of the illegitimate recombination. Underlined letters are the tri-nucleotide repeat shaded in Fig 6A.

C**.** High indel polymorphism at the linker region between CC and NBS domains. Upper case residues are identical or similar to the consensus. Lower case residues are dissimilar. Lower case x in the consensus sequence are non-consensus residues. Bold and underlined residues are part of a repeat.

Fig 1.

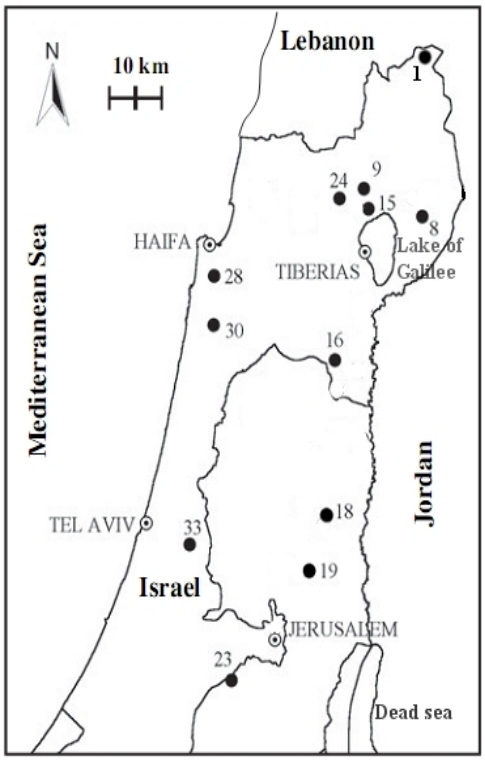
N
10 km
Lebanon
1
24
9
15
8
HAIFA
Lake of
Galilee
TIBERIAS
28
30
16
Mediterranean Sea
Jordan
TEL AVIV
33
18
19
Israel
JERUSALEM
23
Dead sea

Fig 2.

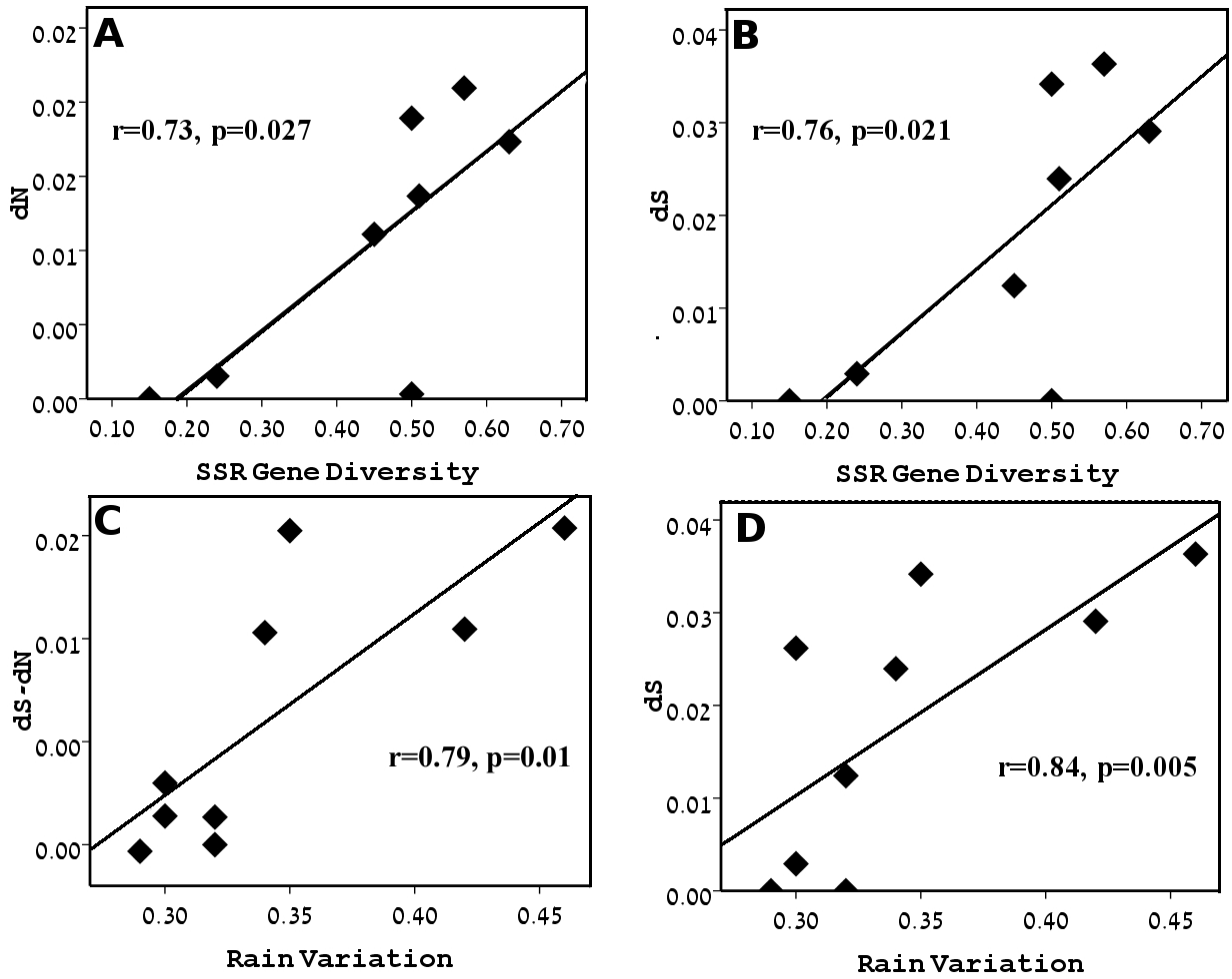
A
r=0.73, p=0.027
dN
SSR Gene Diversity
B
r=0.76, p=0.021
dS
SSR Gene Diversity
C
r=0.79, p=0.01
dS-dN
Rain Variation
D
r=0.84, p=0.005
dS
Rain Variation

Fig 3.

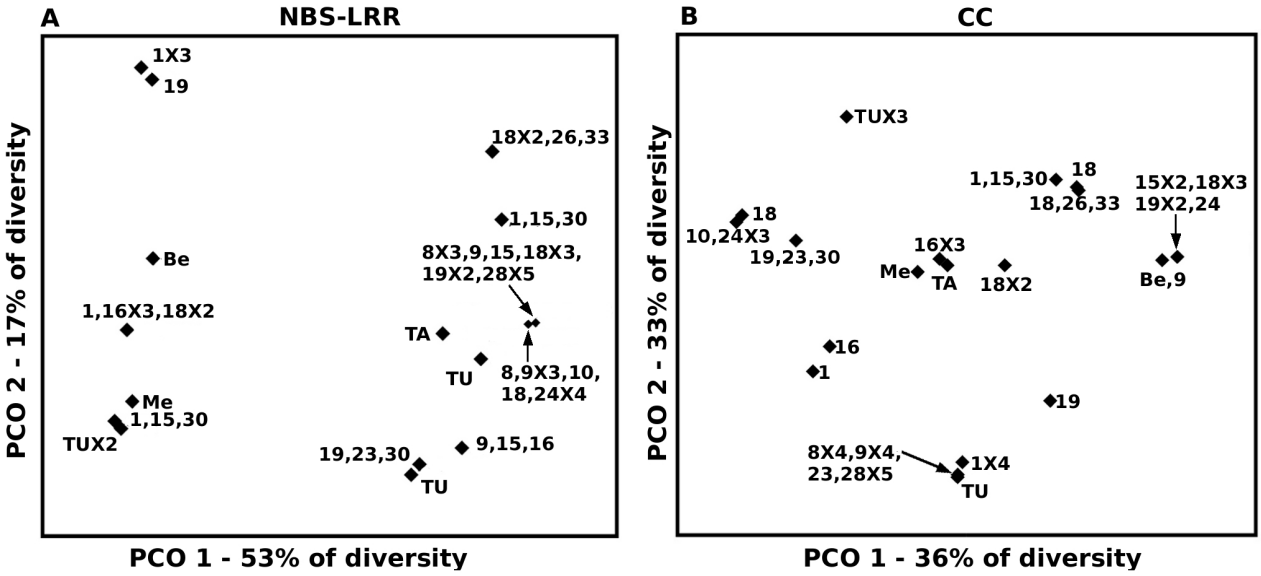
A
NBS-LRR
1X3
19
18X2,26,33
1,15,30
Be
8X3,9,15,18X3,
19X2,28X5
1,16X3,18X2
TA
TU
8,9X3,10,
18,24X4
Me
1,15,30
TUX2
19,23,30
9,15,16
TU
PCO 2 - 17% of diversity
PCO 1 - 53% of diversity
B
CC
TUX3
1,15,30
18
18,26,33
15X2,18X3
19X2,24
18
10,24X3
19,23,30
16X3
Me
TA
18X2
Be,9
16
1
19
8X4,9X4,
23,28X5
1X4
TU
PCO 2 - 33% of diversity
PCO 1 - 36% of diversity

Fig 4.

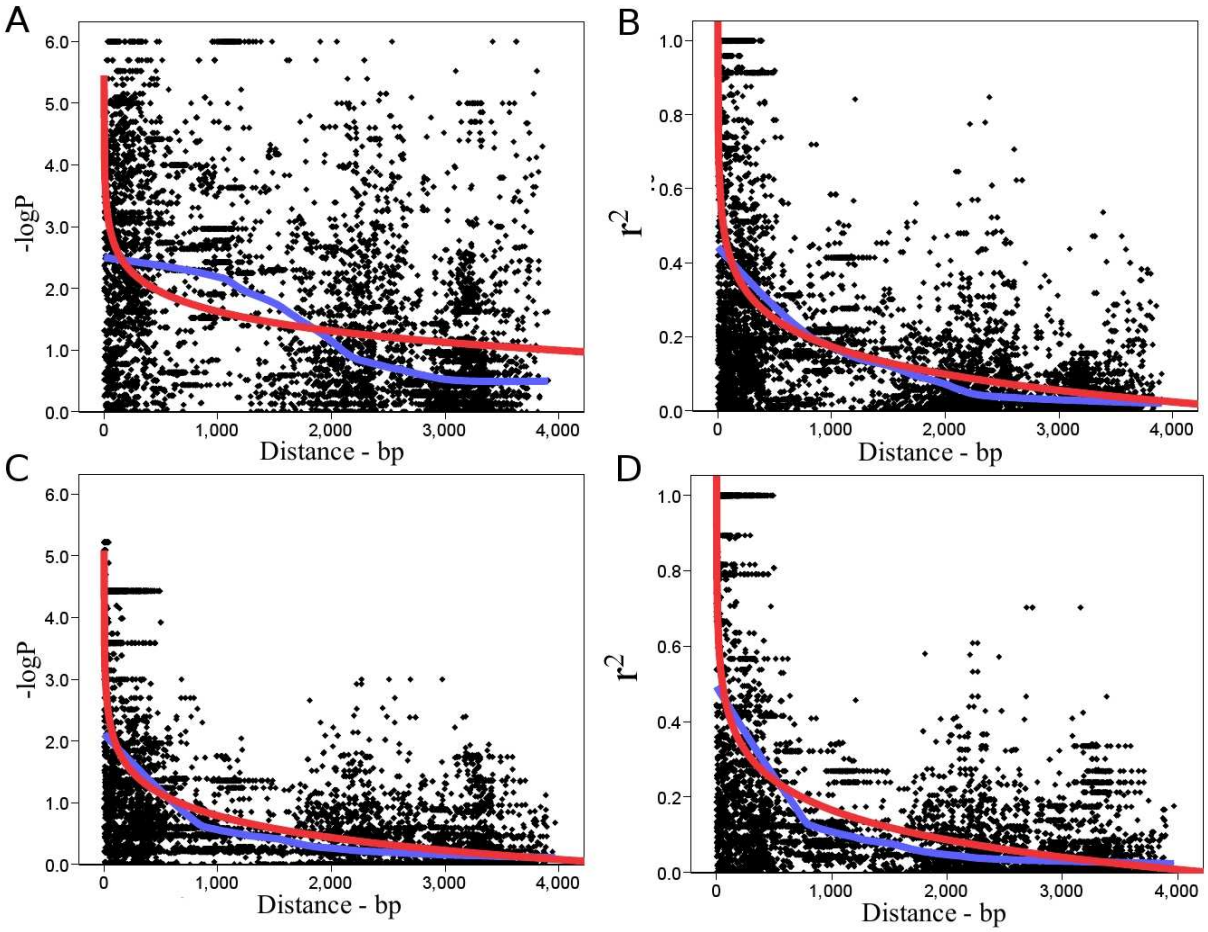
A
B
C
D
-logP
r2
Distance - bp
0
1,000
2,000
3,000
4,000
0.0
1.0
2.0
3.0
4.0
5.0
6.0
0.2
0.4
0.6
0.8

Fig 5.

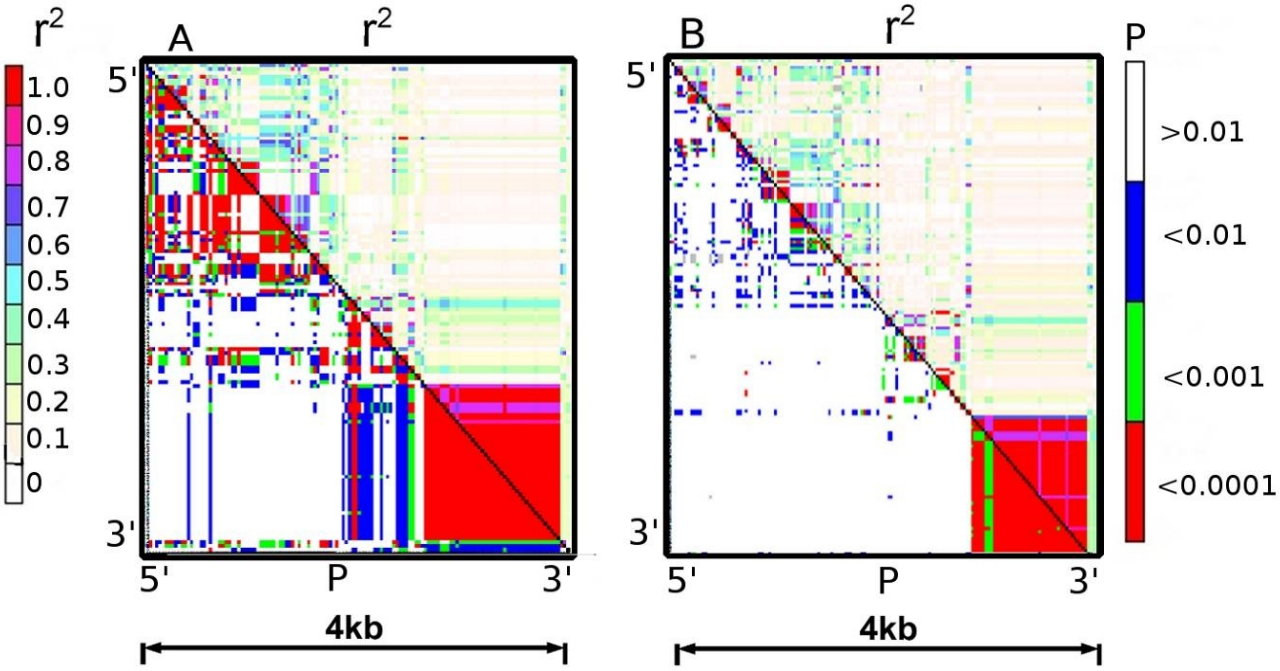
r2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A
r2
5'
3'
5'
P
3'
4kb
B
r2
5'
3'
5'
P
3'
4kb
P
>0.01
<0.01
<0.001
<0.0001

Fig 6.

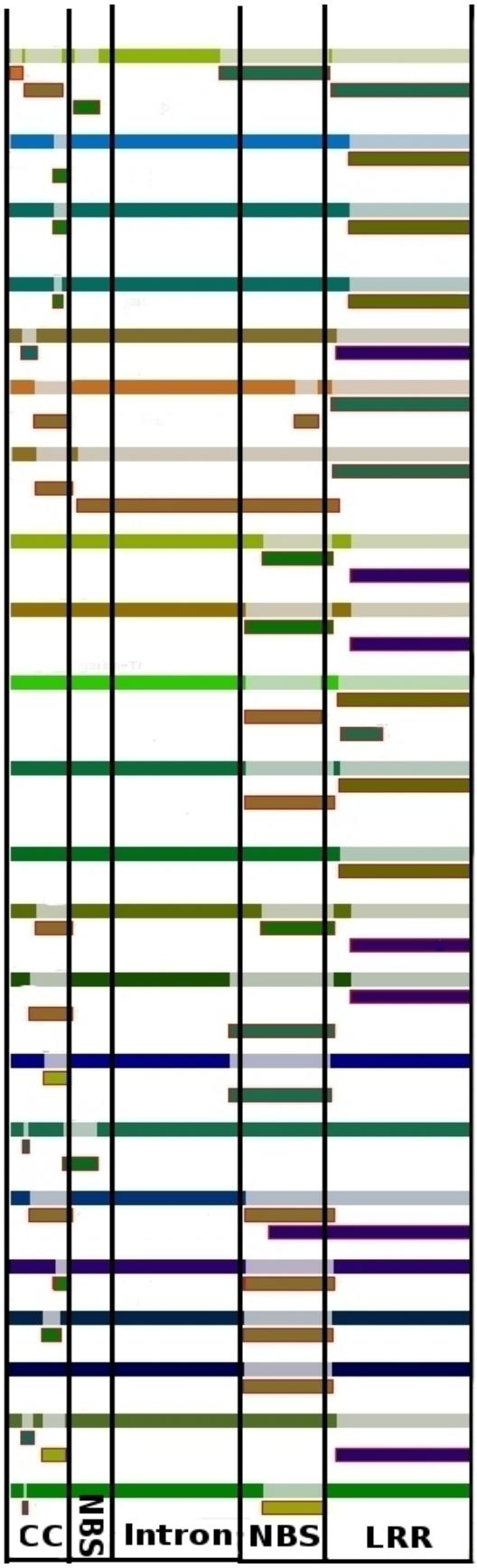

Fig 7.

A.

1471 **1497**

TCGC**CTG**CGCGGCACGACACCCGCTGCAGT**CTG**

TCGC----------CTG

B.

1448 **1480** **1512**

ACCACACTCGCTGTAGTCCTCGCCTGCGCGGC**ACGAC**ACCCGCTGCAGTCTGCGTCCGTGCGAC**ACGAC**

ACGACACCCGCTGCAGTCTGCGTCCGTGCGAC**ACGAC**

1480 **1512**

C.

Consensus	EAHxxxxxxxxERY-xVEDxxxxxxxxxVCP
1_59	Kvs-----nrcERY-wVED---------VCP
30_4	Evg-----**nryERY-rV**v**Nrceryrv**edVCP
Be_1	EAH--------DRY-mVDy---------VyP
18_1	EvpyqiileahDRY-rVED---------VCP
Me_2	EAH--------DR-smVDy---------VCP
18_25	EAH--------E-Y--------------VCP
13_53	EAH-----drcERYcrVED---------VCP
Consensus	EAHxxxxxxxxERY-xVEDxxxxxxxxxVCP